Published in final edited form as:

Mol Ecol. 2007 December ; 16(24): 5246–5258.

Selective sweeps in the homoploid hybrid species *Helianthus deserticola*: evolution in concert across populations and across origins

BRIANA L. GROSS^{*,†}, KATHRYN G. TURNER^{*,§}, and LOREN H. RIESEBERG^{*,†}

^{*} Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

[†] Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4

Abstract

The evolution of different populations within a species in response to selective pressures can potentially happen in three different ways. It can occur in parallel, where similar changes occur independently in each population in response to selection; in concert, where the spread of an adaptive mutation across a species' range results in a single allele fixing in each population; or populations can diverge in response to local selective pressures. We explored these possibilities in populations of the homoploid hybrid species *Helianthus deserticola* relative to its parental species *Helianthus annuus* and *Helianthus petiolaris* using an analysis of variation in 96 expressed sequence tag-based microsatellites. A total of nine loci showed evidence consistent with recent selection at either the species or population level, although two of these genes were discarded because the apparent sweep did not occur relative to the parent from which the locus was derived. Between one and five loci showed a putative sweep across the entire species range with the same microsatellite allele fixed in each population. This pattern is consistent with evolution in concert despite geographical isolation and potential independent origins of the populations. Only one population of *H. deserticola* showed candidate sweeps that were unique compared to the rest of the species, and this population has also potentially experienced recent admixture with the parental species.

Keywords

Helianthus; hybrid species; *lnRH*; microsatellites; parallel evolution; selective sweep

The questions of how species originate and how populations of a species maintain coherence or diverge is of great interest to evolutionary biologists (Ehrlich & Raven 1969; Morjan & Rieseberg 2004). One way to approach this question is to pursue systems in which speciation has occurred in parallel, which greatly increases the chance of identifying causative factors that are constant across all origins (speciation events resulting in a single taxon) (Schluter 2001; Nosil *et al.* 2002; McKinnon *et al.* 2004). In cases where there is gene flow among origins, taxa with multiple origins can also answer questions about how species identity is maintained—they offer what is perhaps the most difficult scenario of integrating populations that may not be identical by descent.

The factors contributing to parallel speciation events can be investigated through the use of experimental methods that recreate the putative conditions leading to the origin of a species.

Correspondence: Briana L. Gross, Washington University, One Brookings Drive, Department of Biology Box 1137, St. Louis, MO 63130, USA. Fax: (314) 935 5125; E-mail: brianagross@wustl.edu.

[‡]Present address: Department of Biology, Washington University, St. Louis, MO 63130, USA

[§]Present address: College of Natural Sciences, University of Texas at Austin, Austin, TX 78712, USA

This approach has been especially fruitful for polyploid (Song *et al.* 1995; Schranz & Osborn 2000; Hegarty *et al.* 2006) and homoploid (Lexer *et al.* 2003; Gross *et al.* 2004; Ludwig *et al.* 2004) hybrid species, where it is possible to recreate the ancestral genotype by hybridizing the two parents. The recent evolution of species with multiple origins is also interesting and can be informative about the circumstances surrounding the formation and early obstacles faced by a new lineage. One question that applies to both the early and current evolution of species with multiple origins is whether, and with what frequency adaptive alleles have spread across the species range. The early spread of adaptive alleles across multiple origins of a species would be one key to integrating different origins. On the other hand, it is also possible that continuous parallel evolution of each origin could generate a phenotypically coherent species without ongoing gene flow.

The homoploid hybrid species *Helianthus deserticola* is an excellent study system in which to investigate the spread of adaptive alleles across a species range and across origins. *Helianthus deserticola* is one of three hybrid sunflower species (including *H. anomalus* and *H. paradoxus*) that are the product of hybridization between *H. annuus* and *H. petiolaris* (Rieseberg 1991), and while the species share the same chromosome number ($n = 17$) and an annual, outcrossing life history, they differ greatly in habitat. Specifically, *H. annuus* is widely distributed throughout the central and western USA, and tends to be found in clay-based soils. *Helianthus petiolaris* has a smaller range, occurring mainly in the central USA and as far west as Utah, and tolerates sandier, dryer soils than does *H. annuus*. The hybrid species *H. deserticola* is a xerophyte found in loose, sandy soils of the Great Basin Desert. It has a restricted range and is found only in small populations in western Nevada, west central Utah, and along the border of Utah and Arizona (Heiser 1947; Heiser *et al.* 1969; Rogers *et al.* 1982). Moreover, *H. deserticola* may have as many as three different origins, based on evidence from chloroplast DNA (cpDNA), microsatellites, and population interfertility (Gross *et al.* 2003). The extreme, divergent habitat of the hybrid compared to the parental species may have been important in promoting the multiple origins, because it could have resulted in strong, parallel, habitat-based selective pressures across the species range.

How has *H. deserticola* adapted to its extreme habitat? The production of transgressive phenotypes in hybrids seems to have played a major role (Gross *et al.* 2004). Presumably, however, there has been continued selection since the origin of the species [between 170 000 and 63 000 years before present (Gross *et al.* 2003)], which would have lead to the fixation of adaptive alleles. The genetic signature of recent adaptations are recorded in the genome in the form of reduced genetic variation, and such reductions can be detected with the aid of multilocus genomic scans or hitchhiking mapping (Harr *et al.* 2002; Schlotterer 2002; Schlotterer & Dieringer 2005). Genetic variation should be lowest near the target of selection and rise with increasing genetic distance from the selected site(s) (Maynard Smith & Haigh 1974). In this study, we used a multilocus genome scan to test for regions of low variation across all potential origins of the homoploid hybrid species *H. deserticola* relative to its two parental species, *H. annuus* and *H. petiolaris*. The scan was conducted using microsatellites located within expressed sequence tags (ESTs), which means they are closely linked to genes that are expressed in *Helianthus*. We used the *lnRV* and *lnRH* approaches (Schlotterer 2002; Schlotterer & Dieringer 2005) along with distance-based methods (Beaumont & Nichols 1996) to identify a ranked list of candidates for selection from 96 loci. In addition, we evaluated the genomic composition and population structure of *H. deserticola* based on these microsatellites to further explore the origin of the hybrid species.

Materials and methods

Plant material

Achenes were collected from four natural populations each of *H. annuus*, *H. petiolaris*, and *H. deserticola* between June 2000 and July 2004. All populations were collected from the Great Basin Desert within the known range of *H. deserticola* to avoid confounding the adaptation of the parental species to the generally dry habitat of this region with the unique adaptation of *H. deserticola* to the most arid areas (Table 1, Fig. 1). The collections included all three potential origins of *H. deserticola* and covered major geographical divisions within the species range: western Nevada, west central Utah, and the Arizona/Utah border from both the Colorado and Virgin River basins. Seedlings were grown in the Indiana University greenhouse and DNA was extracted from fresh tissue using the QIAGEN DNeasy 96 Plant Kit. DNA of a minimum concentration of 15 ng/μL was extracted from 48 individuals per population, for a total of 192 samples of each species and 576 samples in all.

Microsatellites

Screening employed previously developed primers for microsatellites found in the flanking regions of ESTs. All microsatellites were developed from sequences in the Compositae Genome Project database (www.compgenomics.ucdavis.edu/index.htm). A total of 258 microsatellites were screened on a panel of 12 individuals for amplification in all three species and for polymorphism in at least one of the three species. One hundred and sixteen microsatellites were recovered that matched both of these requirements.

PCR

Fragments were fluorescently labelled during the polymerase chain reaction (PCR) using a three-primer system, wherein a fluorescent dye anneals to the 5' end of the forward primer (Schuelke 2000). PCR reactions were performed in volumes of 10 μL containing 2 ng DNA, 1 unit *Taq* DNA polymerase, 0.02 μM of forward primer, 0.20 μM of reverse primer, 0.20 μM fluorescent dye (6FAM, VIC, NET, or PET), 20 mM Tricine, 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgCl₂, 0.1% Triton X-100, 0.001 mg of molecular-grade bovine serum albumin, and 100 μM of each dNTP. Fragments were amplified using a 'touchdown' PCR protocol, developed to reduce nonspecific primer binding and fragment amplification (Don *et al.* 1991). An initial denaturing cycle of 3 min at 95 °C was followed by 10 touchdown cycles (starting at 58 °C and dropping by 1 °C each cycle to the final annealing temperature of 48 °C) of 30 s at 94 °C, 30 s at the intermediate annealing temperature, and 45 s at 72 °C. These 10 cycles were followed by 29 cycles of 30 s at 94 °C, 30 s at 48 °C, and 45 s at 72 °C, with a final elongation period at 72 °C for 20 min.

Genotyping

Microsatellite fragments were assayed via electrophoresis on an ABI 3730 capillary sequencer (Applied Biosystems). PCR fragments of nonoverlapping size and colour were pooled and diluted 1:1 of the diluted 50 in ddH₂O. Finally, 2 μL PCR product was added to a 10-μL mixture of 8.95 μL L of GenScan-500 Liz Size Standard ddH₂O and 0.05 μL (Applied Biosystems). Samples were denatured at 95 °C for 5 min and snap cooled on ice before loading onto the 3730.

Genotypes were scored using GENEMAPPER (version 3.7) (Applied Biosystems). Peaks were scored automatically based on allelic bins created for each microsatellite and then corrected manually for cases of errors or nonsensical automatic scoring. Microsatellites were rerun if more than 10% of the data was missing, and data from the two runs were compared to ensure primer fidelity. Some microsatellites were discarded at this stage because mispriming in certain

populations or consistently weak amplification prevented the collection of 90% of the data. In a few cases, microsatellites were discarded if there was a large gap in allele sizes indicative of nonstepwise mutational effects. Complete data were collected for 96 loci.

Genomic composition and population structure

The two parental species are extremely divergent at the karyotypic level, and *H. deserticola* is karyotypically unique compared to both parents. The divergent chromosomal architecture of the two parental species prevents regular pairing at meiosis (Chandler *et al.* 1986), and the genomic architecture of *H. deserticola* (and the two other hybrid sunflower species) appears to consist of large blocks of chromosomal material inherited from one parent or the other, at least within the current level of resolution based on comparative mapping (Burke *et al.* 2004; Rieseberg *et al.* 1995, 1996). Note that this is a contrast to genome-wide recombination that might occur upon hybridization of two species with collinear chromosomes, and it means that a given genetic locus (and therefore the alleles at that locus) is likely to be derived from only one of the parental species. To take this unique genomic composition into account, we determined the parental origin of each locus in *H. deserticola* and used this to inform our conclusions about the biological reality of potential selective sweeps.

Genomic composition was evaluated using a hybrid index estimated using maximum likelihood as implemented in the program HINDEX (Buerkle 2005). The hybrid index (values range from 0 to 1, where 0 is pure *H. petiolaris* and 1 is pure *H. annuus*) of each *H. deserticola* individual was evaluated based on the individual's genotype and the frequencies of the alleles at the 96 loci in the two parental species. This provided an overall hybrid index for each population of *H. deserticola* based on the entire genomic scan. In addition, hybrid indices were calculated across *H. deserticola* populations for individual loci so that a parental origin could be assigned to each locus. Results from this second analysis should be approached with caution, because the strength of this method of hybrid index estimation lies in using multiple loci simultaneously, and using only one locus results in wide confidence intervals. However, the results are of interest for other analyses, and several measures were taken to avoid spurious and uninformative assignments. Hybrid index scores were averaged for each population and assigned as '*H. petiolaris*' if the value was less than 0.45, or '*H. annuus*' if the value was greater than 0.55. A locus was assigned to one parent or the other for the entire species only if the assignments for all four populations could be made and were identical. As a check, the hybrid indices of *H. deserticola* populations were recalculated using only loci that were assigned to one parent or the other to ensure that all populations received a hybrid index to match the expected parent. The results were as expected; hybrid index scores calculated using only loci assigned to *H. petiolaris* resulted in *H. deserticola* population scores between 0.1 and 0.003, while hybrid index scores calculated using only loci assigned to *H. annuus* resulted in *H. deserticola* population scores between 0.89 and 1.0.

Phylogenetic analysis of microsatellite frequency was performed as in Gross *et al.* (2003) to re-assess the evidence for multiple origins based on nuclear markers, which would result in a pattern of polyphyly or paraphyly for the hybrid species relative to the parental species. Briefly, neighbour-joining trees were constructed based on Nei's genetic distance and support for nodes was evaluated based on 1000 bootstrap replicates using PHYLIP 3.66. One tree was constructed using all loci and all populations and two trees were constructed based on loci derived from one of the parental species and containing populations of the hybrid species and the appropriate parental species (for these trees, the pooled allele frequencies of the other parent were used as an outgroup). Building trees based on microsatellites derived from only one parent is necessary because independent hybrid lineages can become fixed for the same parental loci (Rieseberg *et al.* 1996), potentially causing populations with unique origins to appear more similar to each other than to any populations of the parental species when all loci are included in the analysis.

Species and population structure was more thoroughly investigated using a Bayesian clustering method as implemented in STRUCTURE (version 2.1) (Pritchard *et al.* 2000; Falush *et al.* 2003). All individuals were analysed using the ‘no admixture’ model, with the value of K (number of groups) allowed to range from 2 to 5. Three independent runs with a burn-in of 20 000 and 100 000 Markov chain Monte Carlo for each value of K were used to evaluate the assignment of individuals to the correct species (or population, when $K > 3$). The analysis was then repeated allowing admixture, to evaluate the genomic composition of *H. deserticola* populations relative to the parents. Output of STRUCTURE was graphically displayed using the program DISTRUCT (Rosenberg 2004).

Identification of outlier loci

lnRV and lnRH—Variance in microsatellite fragment length (repeat number) and expected heterozygosity (H_E , or gene diversity) were calculated using the program MICROSATELLITE ANALYSER (Dieringer & Schlotterer 2003). Global F_{ST} and all possible pairwise F_{ST} values were calculated according to the method of Weir & Cockerham (1984) as implemented in GENEPOP (Raymond & Rousset 1995).

Populations were grouped according to species or population designation to test for outlier loci. Groupings of *H. deserticola* were tested for significant reductions in variance and heterozygosity compared to parental species using the lnRV and lnRH tests developed by Schlotterer (2002) and Schlotterer & Dieringer (2005), respectively. The two statistics were calculated using the following equations:

$$\ln RV = \ln \left(\frac{V_{pop1}}{V_{pop2}} \right)$$

and

$$\ln RH = \ln \left(\left(\left(\frac{1}{1 - H_{pop1}} \right)^2 - 1 \right) / \left(\left(\frac{1}{1 - H_{pop2}} \right)^2 - 1 \right) \right),$$

where R = ratio, V = variance in microsatellite repeat number, and H = heterozygosity or gene diversity.

Note that using the *ratio* of the variance or heterozygosity present in two groups accounts for differences in mutation rate among loci, thus allowing accurate comparisons. Computer simulations suggest the distribution of lnRV and lnRH values will be statistically indistinguishable from the normal distribution over a variety of demographic conditions, and the outlier loci for each comparison can be assigned a P -value based on this distribution. The P -values represent the probability of observing a given lnRV or lnRH value under the null hypothesis of neutral evolution (Harr *et al.* 2002; Schlotterer 2002; Schlotterer & Dieringer 2005). For simplicity, values of lnRV and lnRH were standardized to a mean of 0 and a standard deviation of 1 to assign significance.

Comparisons were made between groups of the same composition— either species to species or population to population. *Helianthus deserticola* was compared to only one parent at a time to avoid artificially inflating the variation of the parents and generating false positives. *Helianthus deserticola* was first compared to *H. annuus* and *H. petiolaris* at the species level. In this case, there are two tests that must be significant in order for a locus to be considered a candidate for selection: lnRV and lnRH for *H. annuus* or lnRV and lnRH for *H. petiolaris*. All populations of each species were initially included in these comparisons. However, results from STRUCTURE (see Results) indicated that one population of *H. deserticola*, DES1, appeared to differ from the other three populations in having a greater abundance of identifiably

parental genetic material. Thus, the analysis was repeated using only the remaining three populations of *H. deserticola* to represent the species group. Finally, each of the four populations of *H. deserticola* was compared to each of the eight parental populations. In this case, there are eight tests that must be significant in order for a locus to be considered a candidate for selection in *H. deserticola*: $\ln RV$ and $\ln RH$ for all four populations of *H. annuus*, or $\ln RV$ and $\ln RH$ for all four populations of *H. petiolaris*.

In this study, and in general, these statistics are useful to identify *candidate* loci, but more detailed sequenced-based tests are required to confirm any putative signature of selection and identify the region subject to selection. The microsatellites may be physically close to or distant from the actual target of selection; in this case, it is actually possible that the ESTs in which the microsatellites are found are the subject to selection. In addition, as for any statistical test, $\ln RV$ and $\ln RH$ are subject to false positives. This is particularly worrisome if a population or species has experienced one or more bottlenecks, which are known to increase the variance in diversity among loci (Wiehe *et al.* 2007). In this scenario, outlier loci might represent the extreme values in such a distribution. Using both tests reduces this possibility, because the two tests depend on different measures of variation at a locus, and, according to Schlotterer & Dieringer (2005), using the tests in combination can reduce the number of false positives by a factor of three. Because we were already using this fairly stringent requirement of two significant results, we used $P = 0.05$ as a cut-off for significance rather than using the more conservative Bonferroni correction. In any case, as noted above, the putative swept loci identified in this study are useful as candidates for further investigation.

Distance-based—Loci showing high levels of F_{ST} relative to other loci for the same population comparisons have generally been assumed to be subject to divergent selection (Lewontin & Krakauer 1973). This assumption can be quantified using distance-based measures of selection to assign significance to differences in F_{ST} . The coalescent-based method of Beaumont & Nichols (1996), as implemented in FDIST2 (www.rubic.rdg.ac.uk/~mab/software.html), was used as an alternate test of selection to supplement the measures of $\ln RV$ and $\ln RH$. A sample dataset of 100 000 loci was generated to match the average F_{ST} of the real dataset. The sample dataset was based on the same number of populations as the number being compared (depending on the desired comparison, either with or without DES1), with 48 individuals in each population to match the actual data. The F_{ST} values of the loci in the real dataset were then compared to the distribution of F_{ST} values for the simulated loci to identify any outlying values (those not contained within an envelope in which 95% of the data points are expected to lie) while controlling for heterozygosity.

Null alleles—Null alleles are recognized as a potential problem for any microsatellite-based study, especially those applying microsatellite primers across species boundaries (e.g. Chapuis & Estoup 2007). In the context of this study, it is possible that *H. deserticola* might appear to have reduced variation at a locus due to the effect of null alleles. We tested for the presences of null alleles in *H. deserticola* at candidate loci using the Hardy–Weinberg exact test as implemented in GENEPOP (Raymond & Rousset 1995). In most cases, *H. deserticola* populations showing a potential selective sweep were fixed for a single allele, so we applied this test to populations that were not significant or not included in the comparison (usually DES1, which was excluded from some species-wide comparisons).

Sequencing

Candidate loci from species level tests were sequenced in a few representative individuals from the three taxa to ensure that the scored fragments were homologous and contained the predicted microsatellite repeat sequence. Within *H. deserticola*, two individuals from each population

were sequenced to ensure that the fixed allele at each locus actually had the same repeat number throughout the species range.

Amplification of sequencing template was identical to amplification of PCR products for genotyping, except that the fluorescent dye was replaced by water and the final annealing temperature of the touchdown PCR program was raised to 52 °C to improve specificity. Unincorporated primers were removed using ExoSAP-IT (USB). Sequencing reactions were performed in volumes of 10 µL containing 25–30 ng template DNA, 0.40 µM primer, 1.5 mM MgCl₂, and 0.5 µL ABI BIGDYE TERMINATOR version 3.1. An initial denaturing cycle of 1 min at 96 °C was followed by 25 cycles of 10 s at 96 °C, 5 s at 50 °C, and 4 min at 60 °C. Sequences were resolved on an ABI 3730 capillary sequencer at the Indiana Molecular Biology Institute.

Results

Genomic composition and population structure

The hybrid index analysis was used to investigate two different aspects of *H. deserticola* genome structure. The first objective was to evaluate the genomic composition of the different populations of *H. deserticola*. All populations of *H. deserticola* showed about equal contributions from *H. annuus* and *H. petiolaris* (Table 2), although there appeared to be a slight trend towards a larger contribution from *H. petiolaris* in the DES1 population in northern Arizona as compared to the larger contribution from *H. annuus* in the DESC population from western Nevada. This trend makes some sense in light of the fact that DESC grouped more closely with ANNReno than with other populations of *H. deserticola* based on Gross *et al.* (2003). Interestingly, DES5, the population of *H. deserticola* from Little Sahara, did not show an unusually large contribution from *H. annuus*, even though it is the only population with *H. annuus* cpDNA.

The second objective was to assign a hybrid index score to individual loci as a way of determining the parental origin of a given locus. Using the criteria of consistent scores across all four populations, only 41 loci could be unambiguously assigned—25 to *H. annuus* and 16 to *H. petiolaris* (Table 2). This bias towards *H. annuus* was also present within each population. However, these latter estimates must be interpreted with caution, because they do not meet the more strict criteria used for assignment across the four populations.

Neighbour-joining trees were constructed using all 96 loci and including all populations, using the 25 loci derived from *H. annuus* and including populations of *H. annuus* and *H. deserticola*, and using the 16 loci derived from *H. petiolaris* and including populations of *H. petiolaris* and *H. deserticola*. *Helianthus deserticola* forms a single, well-supported clade in all neighbour-joining trees irrespective of the loci used in the analysis (Fig. S1).

Although it is somewhat artificial, the STRUCTURE analysis without admixture was used as a rough test of how well the three species matched their expected population and species identities. Runs allowed between two and five different groups ($K = 2-5$), and in every case individuals were assigned correctly to their expected species and population (when a population of 48 was delineated). Species groups matched expectations when $K = 3$, and the generation of more groups when $K = 4$ or 5 never resulted in a group that crossed species boundaries. Replicate runs with the same K assigned different populations as unique groups when K exceeded 3, which seemed to indicate that such divisions were not biologically meaningful. With admixture allowed, individuals were still always assigned to the correct species group, and most individuals showed minimal contributions from other populations. The exception to this was DES1, which showed contributions of identifiably *H. annuus* and *H. petiolaris* genetic

components to a much greater extent than other populations of *H. deserticola*, regardless of the value of K (Fig. 2).

Outlier loci

Average values for $\ln RV$ and $\ln RH$ were fairly close to zero for all species to species comparisons, ranging from 0.024 to -0.476 . As indicated by these values, the variance in repeat number and H_E for *H. deserticola* were not extremely depressed compared *H. annuus* and *H. petiolaris*, although they were lower (Table 3). Average values for the population level comparisons tended to be negative and sometimes much lower than -1 . Distributions were always standardized before significance was assigned to values of $\ln RV$ or $\ln RH$.

Only one locus showed significant $\ln RV$ and $\ln RH$ for *H. deserticola* at the species level, being significant (or nearly so) relative to both parents, when DES1 was included (Table 4). Interestingly, no loci showed strong significance for $\ln RV$ and $\ln RH$ relative to one species but not the other. However, four loci were significant relative to both species when DES1 was removed (including the locus that was significant with all populations), and two loci showed significance relative to *H. petiolaris*. Because the reduction in sample size of the *H. deserticola* species group with the removal of DES1 could have influenced these statistics (resulting in a ratio of 144: 192 instead of equal sizes), the analysis was also conducted after randomly removing 12 individuals from each parental population. Results were similar and there was no dramatic change in significance levels, although the significance of an $\ln RV$ value was reduced from $P < 0.05$ to $P < 0.10$ for one locus, with the same pattern for $\ln RH$ at another locus. Although the small sample sizes greatly reduced the power of the population level tests, four loci in two populations of *H. deserticola* were significant (or nearly so) for $\ln RV$ and $\ln RH$ (Table 4). One locus was also significant at the species level, as would be expected. Three of these population level comparisons were significant compared to only one parent, and one was significant in comparison to both.

The distance-based test for outlier loci performed in FDIST2 was used to confirm the results of the $\ln RV$ and $\ln RH$ tests. Distance-based tests are of interest in this context because they evaluate selection based on different assumptions, but they are also prone to false positives because of sensitivity to demographic history and differences in mutation rates. As might be predicted from these circumstances, many more loci were statistically significant according to the FDIST2 results (at least 12) compared to the $\ln RV/\ln RH$ tests. Thus, we report only results that are also supported by $\ln RV/\ln RH$ tests: in this case, three loci. Measures of global F_{ST} were consistent with the identification of outlier loci using other methods, yielding a higher value for the nine outlier loci (0.497) than the 87 nonsignificant loci (0.369), and the difference in averages of single locus measures for the two groups was marginally significant (Mann–Whitney U test, $Z = -1.879$, $P = 0.060$).

We were able to apply the Hardy–Weinberg exact test to at least one population for each candidate locus, testing a total of 14 populations across the nine loci. The null was rejected for one population at each of three different loci (data not shown). Of these three loci, one was later discarded because it showed a potential sweep relative to a parent from which it was not derived (see discussion). Another locus, significant at the population level, deviated from Hardy–Weinberg equilibrium in only one of three populations tested. Overall, the fact that most populations of *H. deserticola* appear to be in Hardy–Weinberg equilibrium when there is variation indicates that null alleles are not a major problem for the identified candidate loci.

Three loci in this study are known to occur on different chromosomes in *H. annuus* (S. Knapp, personal communication, 2004), the rest are unmapped. It is possible that some of the loci are on the same chromosomes in either *H. petiolaris* or *H. deserticola* given the extreme chromosomal rearrangement of the sunflower species relative to each other (Burke *et al.*

2004); further mapping would be necessary to confirm or discount this possibility. Of the nine candidate loci, only five have a match in GenBank, and only four have a predicted function or expression information (Table 4). Sequencing will be necessary to extend the original read length from the EST library and identify probable orthologs. In all cases, *H. deserticola* or one of its individual populations showed complete or almost complete fixation of a single allele, with a maximum of three alleles—one at high frequency and the other two appearing only once. All of the alleles that are found at high frequencies in *H. deserticola* are also present in the parental species, rather than being unique variants found only in the hybrid species, although the alleles at locus HT0719 and locus HT0611 are at fairly low frequency in the parents (Table S1).

Representative alleles were sequenced for four of the five outlier loci (HT0544, HT0546, HT0611, and HT0854) that showed a potential selective sweep at the species level relative to one or both parental species, but it was not possible to obtain clean sequence data for HT0497 under the same conditions. The sequenced candidate loci were homologous across species boundaries and all contained the microsatellite repeat sequence originally identified in the *H. annuus* EST sequence. The alleles that were fixed in *H. deserticola* for these candidate loci always contained the same number of microsatellite repeats.

Discussion

Genomic composition and population structure of *H. deserticola*

The origin of *H. deserticola* and its relationship to *H. annuus* and *H. petiolaris* has been investigated previously (Gross *et al.* 2003). However, the use of a substantially greater number of molecular markers and new techniques in this paper has revealed some previously unknown patterns that have implications for this study as well as our previous findings and should be reviewed.

The genomic composition of the hybrid species was evaluated in two different ways, first by assigning a hybrid index score to each population based on all 96 loci, and then by assigning a parental origin to each locus independently. The hybrid index score for each population shows that *H. deserticola* seems to contain roughly equal contributions from each parental species with a slight trend towards *H. annuus*. This trend is exaggerated in the assignment of individual loci. A greater number of loci can confidently be assigned an origin in *H. annuus* (25) than can be assigned to *H. petiolaris* (16). While the difference in number of loci is not statistically significant, it is noteworthy that *H. deserticola* is actually phenotypically more similar to *H. petiolaris* than to *H. annuus* (Rosenthal *et al.* 2002) and can be difficult to distinguish from *H. petiolaris* in the field (L.H. Rieseberg and B.L. Gross, personal observation). Moreover, a selection analysis of BC₂ progeny grown in the greenhouse and in the *H. deserticola* habitat indicated that the BC₂ towards *H. petiolaris* resembled the *H. deserticola* phenotype most closely and had high fitness in the field (Gross *et al.* 2004; Rosenthal *et al.* 2005). Thus, it seems paradoxical that the majority of loci should be derived from *H. annuus*. This puzzling pattern was also seen in a previous mapping study based on amplified fragment length polymorphisms (Rieseberg *et al.* 2003) and cannot be explained based on our current experimental work.

The original conclusions about the origins of *H. deserticola* were based on information from cpDNA, microsatellites, and population interfertility (Gross *et al.* 2003). Our previous phylogenetic analysis using eight microsatellite loci derived from *H. annuus* suggested that DESC was more closely related to *H. annuus* than to other populations of *H. deserticola*, and therefore potentially the product of a unique hybrid speciation event. However, the current analysis using 25 microsatellites derived from *H. annuus* do not support this pattern, and, indeed, all possible combinations of loci and populations show *H. deserticola* as a single,

monophyletic group (Fig. S1). This eliminates one of the three possible origins of *H. deserticola*, raising the question of whether multiple origins are really a possibility for this species. The other two origins of the hybrid species were proposed based on the presence of *H. annuus* cpDNA in the equivalent of population DES5 and *H. petiolaris* cpDNA in all other populations of *H. deserticola*. The presence of both types of cpDNA could either be caused by multiple origins or by cytoplasmic introgression between *H. deserticola* and its parents. Results from STRUCTURE, however, do not show admixture between DES5 and the parental species (although this is detected in another population; see below). This strongly supports the possibility that the different cpDNA types in *H. deserticola* result from multiple origins.

The population structure of *H. deserticola*, as revealed by STRUCTURE, also adds novel information that is important to understanding the origin of the species. The patterns of genetic variation in the first study lead to the conclusion that one of the major origins of *H. deserticola* was in the Colorado River Basin, followed by a subsequent westward spread (Gross *et al.* 2003). This was mainly supported by the fact that populations in the Colorado River Basin have the greatest diversity of cpDNA types and contain the type of cpDNA found in most other populations of *H. deserticola*. However, it appears that the current representative from this area, DES1, contains genetic material that is more closely related to parental species, or shows greater admixture with parental species, than any other population of *H. deserticola*, which show almost none (Fig. 2). There are two possible explanations for this phenomenon. One is that DES1 is truly more ancestral and contains alleles that closely match those found in the parental species because it has retained a greater diversity of alleles than the derived populations after the westward spread. The other is that DES1 has hybridized with parental species in the area and has thus gained new parental alleles (and cpDNA in the previous study) that are actually more recently acquired than any of the alleles in other populations of *H. deserticola*. The second explanation seems more likely based on the fact that loci in all populations of *H. deserticola* can be assigned a parental origin, but only DES1 contains alleles with such close identity to those of parental populations for admixture to be detected.

Species and selective sweeps

The question of how or whether species originate and evolve in parallel or in concert is an important one and can be addressed in many ways. This study provides an interesting perspective on the question by introducing the case of multiple origins, adding another layer of complexity to the phenomena of local divergence and species coherence. The homoploid hybrid species *H. deserticola* is thought to have originated between 170 000 and 63 000 years before present, which is roughly equivalent to the same number of generations in this annual species (Gross *et al.* 2003). Given that the average microsatellite mutation rate in plants has been estimated to be fairly high [10^{-4} per generation (Diwan & Cregan 1997; Vigouroux *et al.* 2002)], the putative selective sweeps identified in this paper are unlikely to date from the origin of *H. deserticola* and may be quite recent (Wiehe 1998). This is a weakness of the *lnRV* and *lnRH* tests in the context of understanding the origin of species, as their power decreases significantly with the amount of time elapsed since selection (Schlotterer 2002; Schlotterer & Dieringer 2005). However, the evolutionary dynamics of recent selective sweeps are presumably similar to ancient sweeps that have occurred throughout the history of the *H. deserticola* lineage. Therefore, understanding whether these apparent sweeps are geographically localized or occur across the range of the hybrid species can help us understand both the early evolution of *H. deserticola* following multiple origins and how it is responding to current selective pressures.

How does a species, particularly one with multiple origins, respond to selective pressures? The results in this study suggest that there can be a concerted response despite both geographical isolation and potentially different origins because some loci showed putative selective sweeps

across the species range with the same allele fixed in each population. One locus shows a pattern consistent with a selective sweep across all four populations of *H. deserticola* compared to both parental species. When DES1 (which may be experiencing introgression from the parental species) was excluded from the analysis, three additional loci appear to show a selective sweep relative to both parental species in the three remaining populations. These three populations contain both of the strongest candidates for independent origins based on our previous study and current findings: DES5 (the only population with *H. annuus* cpDNA) and DESC and DES4 (both populations with *H. petiolaris* cpDNA that is found across the range of *H. deserticola*).

Along with putative selective sweeps that occurred relative to both parents, there were two that occurred relative to only one parent at the species level. Interpreting the relevance of this type of putative selective sweep requires consideration of the nature of a homoploid hybrid species. We considered such candidate loci to be biologically meaningful only if the putative sweep occurred relative to the parental species from which the locus was derived. That is, if a locus was derived from *H. annuus* and showed a sweep in *H. deserticola* relative to that parent, it likely represents a change in allele frequency in response to selection. If, on the other hand, that locus was derived from *H. petiolaris*, the detected sweep might be more likely to reflect differences in allele frequencies between the two parental species, and the declaration of a selective sweep would be spurious. For example, HT0495 is derived from *H. annuus*, but shows a selective sweep at the species level relative to *H. petiolaris*. This is likely explained by the fact that the variance in repeat number is highest for *H. petiolaris* (0.79), lower for *H. annuus* (0.234), and lowest for *H. deserticola* (0.112). Thus, *H. deserticola* appears swept relative to *H. petiolaris*, but in actuality it simply has a slightly restricted set of the small number of alleles present in *H. annuus*. Locus HT0611, however, is both derived from *H. petiolaris* and shows a putative sweep at the species level relative to this parent, making it a meaningful candidate locus for further study.

In total, there is one candidate locus for a selective sweep at the species level when all the *H. deserticola* populations are included (HT0854), and four more candidates that are significant when DES1 is excluded (HT0497, HT0544, HT0546, and HT0611). Sequence data shows that HT0544, HT0546, HT0611, and HT0854 are homologous across species boundaries. More importantly, it shows that alleles of the same length in different populations of *H. deserticola* have the same repeat number and are virtually identical at the sequence level, confirming that the same alleles are fixed across the species range for these candidate loci.

As mentioned previously, some of these putative sweeps may be false positives, especially given the potential for a bottleneck event at the origin of *H. deserticola* that is suggested by the slightly lower genetic diversity in the hybrid species compared to the parental species. The results from this study should therefore be considered indicative of overall patterns of evolution in the hybrid species, and patterns of evolution at individual loci must be confirmed with sequence data. However, even if a few of the five putative species-wide sweeps are eliminated, it appears that populations of *H. deserticola* can evolve in concert across the species range.

The fact that some putative sweeps appear to be species-wide suggests that *H. deserticola* might have responded to similar selective pressures via the spread of a single adaptive allele across its range. Studies of microsatellite variation in the parental species *H. annuus* and *H. petiolaris* also suggest that the geographical extent of selective sweeps can extend across the sampled range for a species and lead to the spread of the same allele across all populations (Kane & Rieseberg 2007; Yatabe *et al.* 2007). More importantly, this study is consistent with an analysis of the homoploid hybrid species *H. paradoxus*, where microsatellites linked to survivorship quantitative trait loci in a salt marsh environment showed statistically significant reductions in diversity compared to parental species, and the same alleles were swept across

all populations for some loci (Edelist *et al.* 2006). Although *H. paradoxus* likely has a single origin, rather than multiple origins like *H. deserticola*, both species have restricted ranges consisting of geographically isolated populations, so the spread of adaptive alleles would require long-distance dispersal in either case. This potential for gene flow and consequently identical responses to the same selective pressures may enable the genetic integration of a species following parallel ecological speciation, as has been proposed for *H. deserticola*. It is possible that the extreme desert habitat exerted strong, equivalent selective pressure on all origins of *H. deserticola*, facilitating the spread of identical, adaptive alleles across its range both early in the evolution of the species (following multiple origins) and continuously to the present.

Up to four loci appear to have experienced local selective sweeps at the population level compared to one or both parents despite the reduced power of the population level comparisons. One of these, HT0732, can be discarded because it appears to be under selection relative to *H. annuus*, but is derived from *H. petiolaris*. Another locus, HT0854, was also significant at the species level, so it is not surprising to see it swept at the population level as well. The other two loci, HT0572 and HT0719, are both interesting because they show potential sweeps in DES1, the population that differs from other species of *H. deserticola* in showing greater admixture with parental populations. Although it is only speculative, this does raise the intriguing possibility that the hybrid species, *H. deserticola*, is experiencing adaptive introgression from parental species. Investigating this possibility would require a detailed sampling of *Helianthus* populations in the area near DES1 to document allele frequencies and perhaps pinpoint the origin of the swept alleles.

As stated earlier, outlier loci identified in this paper are to be considered candidates for further study, not confirmed targets of selection. Because these loci are located within expressed sequences, it is possible that those expressed genes are actually under selection. However, it is also possible that these loci are closely linked to a different gene under selection or perhaps in a fairly large region with reduced variation. Linkage disequilibrium resulting from selective sweeps has been shown to extend for up to 250 kb in some cases (Olsen *et al.* 2006), but it also appears that linkage disequilibrium is an order of magnitude lower in wild sunflowers (Liu & Burke 2006), which suggests that the target of selection could be quite close.

Conclusions and future directions

In summary, it appears that populations of *H. deserticola* are evolving in concert across the species range, despite both geographical isolation and potential multiple origins. Although the putative selective sweeps documented in this study are likely recent, the results have implications for how the species might have achieved genetic integration after initially unique origins and parallel evolution under similar selective pressures. The two putative sweeps that appear to have occurred at the population level are both in DES1, a population of *H. deserticola* that has likely experienced recent admixture with the parental species, suggesting the potential for adaptive introgression.

The putative swept loci documented in this study will serve as candidates for sequence-based analysis required to confirm or reject the possibility that the associated genes are actually the targets of selection. Sequence analysis may help to clarify the origin of adaptive alleles in *H. deserticola*—the fact that all microsatellite alleles fixed in *H. deserticola* are also present in the parental species (Table S1) suggests that these are all extant alleles that have become advantageous in the context of the hybrid genome, but it is possible that novel mutations might also play a role. In addition, deeper sampling and sequence analysis is required to elucidate the origin of the alleles that appear to have been swept to fixation in DES1 in relation to the parental species and to quantify the magnitude of potential admixture in this and surrounding populations of *H. deserticola*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Special thanks to N. C. Kane for his assistance with data analysis. Thanks also to J. L. Strasburg for comments on an earlier version of this manuscript and advice on population genetics and sequencing. The comments of two anonymous reviewers greatly improved the quality of the manuscript. Funding for this project was provided by the National Science Foundation Predoctoral Fellowship and the Indiana University Department of Biology Floyd/Ogg Final Year Fellowship to BLG and National Institutes of Health grant (GM059065) to LHR.

Biography

Briana L. Gross studied the origin and evolution of the homoploid hybrid species *H. deserticola* for her PhD, and this paper represents one of her dissertation projects. Kathryn G. Turner is interested in the evolutionary ecology of invasive plants. All of the work was conducted in the laboratory of Loren H. Rieseberg, who studies the genetics of hybridization and speciation in plants.

References

- Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 1996;263:1619–1626.
- Buerkle CA. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* 2005;5:684–687.
- Burke JM, Lai Z, Salmaso M, et al. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 2004;167:449–457. [PubMed: 15166168]
- Chandler J, Jan C, Beard B. Chromosomal differentiation among the annual *Helianthus* species. *Systematic Botany* 1986;11:353–371.
- Chapuis M-P, Estoup A. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 2007;24:621–631. [PubMed: 17150975]
- Dieringer D, Schlotterer C. MICROSATELLITE ANALYSER: a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 2003;3:167–169.
- Diwan N, Cregan PB. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theoretical and Applied Genetics* 1997;95:723–733.
- Don R, Cox P, Wainwright B, Baker K, Mattick J. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 1991;19:4008–4008. [PubMed: 1861999]
- Edelist C, Lexer C, Dillmann C, Sicard D, Rieseberg LH. Microsatellite signature of ecological selection for salt tolerance in a wild sunflower hybrid species, *Helianthus paradoxus*. *Molecular Ecology* 2006;15:4623–4634. [PubMed: 17107488]
- Ehrlich PR, Raven PH. Differentiation of populations. *Science* 1969;165:1228–1232. [PubMed: 5803535]
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003;164:1567–1587. [PubMed: 12930761]
- Gross BL, Kane NC, Lexer C, et al. Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *American Naturalist* 2004;164:145–156.
- Gross BL, Schwarzbach AE, Rieseberg LH. Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *American Journal of Botany* 2003;90:1708–1719.
- Harr B, Kauer M, Schlotterer C. Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99:12949–12954. [PubMed: 12351680]
- Hegarty MJ, Barker GL, Wilson ID, et al. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology* 2006;16:1652–1659. [PubMed: 16920628]

- Heiser C. Hybridization between the sunflower species *Helianthus annuus* and *H. petiolaris*. *Evolution* 1947;1:249–262.
- Heiser C, Smith D, Clevenger S, Martin W. The North American sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club* 1969;22:1–218.
- Kane NC, Rieseberg LH. Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics* 2007;175:1823–1834. [PubMed: 17237516]
- Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 1973;74:175–195. [PubMed: 4711903]
- Lexer C, Welch ME, Raymond O, Rieseberg LH. The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): selection on transgressive characters in a novel hybrid habitat. *Evolution* 2003;57:1989–2000. [PubMed: 14575321]
- Liu A, Burke JM. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 2006;173:321–330. [PubMed: 16322511]
- Ludwig F, Rosenthal DM, Johnston JA, et al. Selection on leaf ecophysiological traits in a desert hybrid *Helianthus* species and early-generation hybrids. *Evolution* 2004;58:2682–2692. [PubMed: 15696747]
- Maynard Smith J, Haigh J. Hitch-hiking effect of a favorable gene. *Genetical Research* 1974;23:23–35. [PubMed: 4407212]
- McKinnon JS, Mori S, Blackman BK, et al. Evidence for ecology's role speciation. *Nature* 2004;429:294–298. [PubMed: 15152252]
- Morjan CL, Rieseberg LH. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* 2004;13:1341–1356. [PubMed: 15140081]
- Nosil P, Crespi B, Sandoval C. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* 2002;417:440–443. [PubMed: 12024213]
- Olsen KM, Caicedo AL, Polato N, et al. Selection under domestication: Evidence for a sweep in the rice *Waxy* genomic region. *Genetics* 2006;173:975–983. [PubMed: 16547098]
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–959. [PubMed: 10835412]
- Raymond M, Rousset R. GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. *Journal of Heredity* 1995;86:248–249.
- Rieseberg LH. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* 1991;78:1218–1237.
- Rieseberg LH, Raymond O, Rosenthal DM, et al. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 2003;301:1211–1216. [PubMed: 12907807]
- Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM. Role of gene interaction in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 1996;272:741–745. [PubMed: 8662570]
- Rieseberg LH, Van Fossen C, Desrochers AM. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 1995;375:313–316.
- Rogers, C.; Thompson, T.; Seiler, GJ. Sunflower Species of the United States. National Sunflower Association; Bismarck, North Dakota: 1982.
- Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 2004;4:137–138.
- Rosenthal DM, Rieseberg LH, Donovan LA. Re-creating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: Three examples using wild sunflowers. *American Naturalist* 2005;166:26–41.
- Rosenthal DM, Schwarzbach AE, Donovan LA, Raymond O, Rieseberg LH. Phenotypic differentiation between three ancient hybrid taxa and their parental species. *International Journal of Plant Sciences* 2002;163:387–398.
- Schlotterer C. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* 2002;160:753–763. [PubMed: 11861576]

- Schlotterer, C.; Dieringer, D. A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: Nurminsky, D., editor. *Selective Sweep*. Springer; New York: 2005. p. 1-10.
- Schluter D. Ecology and the origin of species. *Trends in Ecology and Evolution* 2001;16:372–380. [PubMed: 11403870]
- Schranz ME, Osborn TC. Novel flowering time variation in the resynthesized polyploid *Brassica napus*. *Journal of Heredity* 2000;91:242–246. [PubMed: 10833052]
- Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 2000;18:233–234.
- Song K, Lu P, Tang K, Osborn TC. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences of the United States of America* 1995;92:7719–7723. [PubMed: 7644483]
- Vigouroux Y, Jaqueth JS, Matsuoka Y, et al. Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution* 2002;19:1251–1260. [PubMed: 12140237]
- Weir BS, Cockerham CC. Estimating *F*-statistics for the analysis of population structure. *Evolution* 1984;38:1358–1370.
- Wiehe T. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. *Theoretical Population Biology* 1998;53:272–283. [PubMed: 9679322]
- Wiehe T, Nolte V, Zivkovic D, Schlotterer C. Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics* 2007;175:207–218. [PubMed: 17057237]
- Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* 2007;175:1883–1893. [PubMed: 17277373]

Fig. 1.
Range map showing locations of *Helianthus* populations from Utah, Arizona, and Nevada, USA, used in this study.

Fig. 2.

Results of STRUCTURE analysis for *H. annuus*, *H. deserticola*, and *H. petiolaris* with admixture allowed and $K = 3$. Each individual's genome is represented by a thin vertical line partitioned into coloured segments in proportion to the estimated ancestry in each source cluster. Populations are separated by black lines, with names below. Results were similar for all runs of STRUCTURE with K allowed to range from $K = 2$ to $K = 5$.

Table 1

Population codes, locality, sample size, and codes from equivalent populations used in Gross *et al.* (2003). The 2003 codes are for either the same collections (codes match current codes) or for different collections from the same area (codes are different from current codes)

Taxon/Population	Code	Location (Latitude, Longitude)	N	2003 Code
<i>H. annuus</i>				
Rieseberg 1307	ANN2	Mm 381 on Arizona route 160, W of Kayenta, Navajo Co., AZ (36.66 N, 110.41 W)	48	—
Rieseberg 1310	ANN5	4.8 km S of Toquerville exit of I-15 on Old Hwy 91, Washington Co., UT (37.26 N, 113.28 W)	48	ANN5
Rieseberg 1312	ANN7	3.5 km W of route 6 on road to Little Sahara, Juab Co., UT (39.72 N, 112.23 W)	48	—
Rieseberg Reno	ANNReno	48.3 km E of Reno, Wahoe Co., NV (39.63 N, 119.18 W)	48	ANNReno
<i>H. deserticola</i>				
Rieseberg 1317	DES1	Intersection of 89 and Lake Powell Boulevard on NE corner; just N of Page, Coconino Co., AZ (36.93 N, 111.48 W)	48	DES1274,
Rieseberg 1320	DES4	About 1.6 km S of Toquerville exit off I-15 on frontage road, Washington Co., UT (37.29 N, 113.28 W)	48	DES1270
Rieseberg 1321	DES5	Little Sahara Recreation area, 6.8 km from visitors centre, Juab Co., UT (39.69 N, 112.33 W)	48	DES1265
Rieseberg DesC	DESC	7.7 km N of Hwy 50 on Soda Lake Road, Churchill Co., NV (39.55 N, 118.86 W)	48	DESC
<i>H. petiolaris</i>				
Rieseberg 1322	PET1	Between mm 110 and 11 on route 95; 11.9 km E of intersection of 95 and 261, San Juan Co., UT (37.55 N, 109.79 W)	48	—
Rieseberg 1323	PET2	About 5.6 km N of Arizona/Utah border, Route 163, S of Mexican Hat, San Juan Co., UT (37.03 N, 110.13 W)	48	—
Rieseberg 1325	PET4	On volcanic substrate 0.8 km E of exit 95 (route 20) off I-15, Iron Co., UT (38.08 N, 112.68 W)	48	PET1271
Gross 16/2004	PET16/04	On red substrate, W side of Hwy 89 at mm 60, S of Kanab, Kane Co., UT (37.03 N, 112.48 W)	48	—

Hybrid index scores (standard errors), upper and lower bounds, and parental origin of loci in *H. deserticola*. The hybrid index for each population or locus is based on the allele frequencies in the hybrid species and the two parental species. Hybrid index scores can range from 0 (*H. petiolaris*) to 1 (*H. annuus*). Loci for each population were assigned to *H. petiolaris* if the hybrid index was less than 0.45, or to *H. annuus* if the value was greater than 0.55. Loci were assigned a parental origin for *H. deserticola* as a whole if the assignments for all four populations were identical

Table 2

Population	Hybrid index	Confidence intervals		Origin of loci in <i>H. deserticola</i>		Unknown
	Average hybrid index (SE)	Lower	Upper	<i>H. annuus</i>	<i>H. petiolaris</i>	
DES1	0.502 (0.010)	0.386	0.618	49	39	8
DES4	0.518 (0.006)	0.394	0.641	49	38	9
DES5	0.526 (0.007)	0.407	0.643	46	43	7
DESC	0.589 (0.005)	0.474	0.698	51	37	8
<i>H. deserticola</i>	0.534 (0.004)	0.415	0.650	25	16	55

Table 3

Estimates of genetic diversity within populations and species, and of all individuals pooled (Total), including mean number of alleles per locus, mean variance in repeat number, and gene diversity (H_E) with standard error

Population/Species	Mean # alleles/locus	Mean Rep. # variance	Mean H_E (SE)
ANN2	3.9	4.3	0.49 (0.023)
ANN5	4.9	3.7	0.51 (0.024)
ANN7	3.8	2.9	0.45 (0.024)
ANNReno	3.6	3.0	0.43 (0.024)
<i>H. annuus</i>	4.02	3.5	0.47 (0.020)
DES1	5.7	3.3	0.55 (0.024)
DES4	3.2	2.5	0.39 (0.027)
DES5	3	1.6	0.31 (0.026)
DESC	2.6	1.6	0.29 (0.024)
<i>H. deserticola</i>	3.6	2.3	0.39 (0.018)
PET1	4.5	4.3	0.51 (0.023)
PET2	4	2.5	0.45 (0.023)
PET4	4.2	2.7	0.47 (0.026)
PET16/04	3.5	3.0	0.42 (0.023)
<i>H. petiolaris</i>	4.1	3.1	0.46 (0.018)
Total	9.8	6.2	0.69 (0.016)

Table 4
Results of outlier analyses *lnRV*, *lnRH*, and *FDIST2* (see text), along with the groups compared, origin of each locus, sequence homology, and linkage group for microsatellites mapped in *H. annuus*. Note that HT0854 is listed twice

Locus	Population	Test for selection			Comparison	Locus origin	Homology	Linkage group
		lnRV	lnRH	FDIST2				
<i>H. deserticola</i> , all populations								
HT0572	DES1	4*	INS, 1, 2 ^{†*}	**	Des Pops: Ann Pops	<i>H. annuus</i>	Unknown protein	—
HT0719	DES1	4*	4*		Des Pops: Ann Pops	<i>H. petiolaris</i>	Cytokinin-repressed protein	17
	DES1	4*	4*		Des Pops: Pet Pops			
HT0732	DES4	INS, 3*	1, 3 ^{†*}		Des Pops: Ann Pops	<i>H. petiolaris</i> [‡]	No hit	11
HT0854	<i>H. deserticola</i>	1, 1 ^{†*}	2*	*	Des: Ann and Pet	<i>H. annuus</i>	No hit	13
	DES4	4*	3, 1 ^{†*}	—	Des Pops: Ann Pops			
<i>H. deserticola</i> , without DES1								
HT0495	<i>H. deserticola</i>	1*	1*		Des: Pet	<i>H. annuus</i> [‡]	No hit	—
HT0497	<i>H. deserticola</i>	2*	2*		Des: Ann and Pet	<i>H. annuus</i>	Cytokinin-inducible gene	—
HT0544	<i>H. deserticola</i>	2*	2*		Des: Ann and Pet	<i>H. annuus</i>	Peptidylprolyl isomerase	—
HT0546	<i>H. deserticola</i>	2*	2*		Des: Ann and Pet	<i>H. annuus</i>	No hit	—
HT0611	<i>H. deserticola</i>	1*	1*	**	Des: Pet	<i>H. petiolaris</i>	Putative protein phosphatase	—
HT0854	<i>H. deserticola</i>	2**	1, 1 ^{†*}	*	Des: Ann and Pet	<i>H. annuus</i>	No hit	13

[†]*P* < 0.10,
* *P* < 0.05,
** *P* < 0.01.

[‡]These two loci show a selective sweep relative to a species from which the locus did not originate. See text for details.